HUMAN–COMPUTER INTERACTION, 2004, Volume 19, pp. 225–271

Modeling Information Navigation: Implications for Information Architecture

Craig S. Miller
DePaul University

Roger W. Remington
NASA Ames Research Center

ABSTRACT

Previous studies for menu and Web search tasks have suggested differing advice on the optimal number of selections per page. In this article, we examine this discrepancy through the use of a computational model of information navigation that simulates users navigating through a Web site. By varying the quality of the link labels in our simulations, we find that the optimal structure depends on the quality of the labels and are thus able to account for the results in the previous studies. We present additional empirical results to further validate the model and corroborate our findings. Finally we discuss our findings' implications for the information architecture of Web sites.

Craig Miller is a computer scientist with interests in cognitive modeling and human–computer interaction; he is an Associate Professor in the School of Computer Science, Information Systems and Telecommunications at DePaul University. Roger Remington is an experimental psychologist with interests in visual attention and cognitive modeling; he is a Research Psychologist at NASA Ames Research Center.

CONTENTS

1. INTRODUCTION

The World Wide Web continues to revolutionize how people obtain information, buy products, and conduct business transactions. Yet many companies and organizations struggle to design Web sites that customers can easily navigate to find information or products. Consequently, the identification of factors that affect the usability of the World Wide Web has become increasingly important. Although many of these factors concern the graphical layout of each page in a Web site, the structure of linked pages, often called the site's information architecture, plays a decisive role in the site's usability. The importance of information architecture is attested by the large number of books and articles offering advice on how to best structure information in a Web site (e.g., Rosenfeld & Morville, 1998; Larson & Czerwinski, 1998; Shneiderman, 1998).

Our effort focuses on understanding how a site's information architecture impacts a user's ability to effectively find content in a linked information structure such as a Web site. There have already been a number of empirical studies that evaluate a variety of hierarchical structures in terms of the fastest search times. Most studies have involved menu selection tasks (see Norman, 1991, for a review) but a few have involved Web navigation (e.g., Larson & Czerwinksi, 1998).

Assuming unordered lists of selections, empirical results of menu search experiments consistently favor structures with approximately eight selections per page (Norman, 1991). Structures with as many as eight selections per page produce faster search results than deeper structures with fewer selections per page (Kiger, 1984; D. P. Miller, 1981; Snowberry, Parkinson, & Sisson, 1999); broader structures with more than eight selections per page produce slower search times (D. P. Miller, 1981; Snowberry et al., 1999) unless the pages have naturally organized selections in numeric or alphabetical order (Landauer & Nachbar, 1985) or were naturally grouped in categories (Snowberry et al., 1999). These empirical results are corroborated by a theoretical analysis of hierarchical structures, which suggests that the optimal number of selections per page ranges from 4 to 13, assuming a linear self-terminating search in each page and a reasonable range of reading and key-press times (Lee & MacGregor, 1985).

Despite the apparent similarity of menu selection and Web navigation, results from a study using Web pages appear to be at odds with the conclusions drawn from the menu selection studies. Larson and Czerwinski (1998) examined user search times in Web pages of differing hierarchical depth. In contrast to the results from the menu selection studies, they found that users took significantly longer to find items in a three-tiered, 8-links-per-page (8 × 8 × 8) structure than in comparable two-tiered structures with 16 and 32 links per page (16 × 32 and 32 × 16).

In this article, we examine the apparent discrepancy between the results from menu selection studies and the result from the Web navigation study by Larson and Czerwinski (1998). Some evidence suggests that this discrepancy may be due to the quality of the labels. D. P. Miller (1981) reported that selection errors occurred less than 1% of the time for the 8 × 8 structure used in his menu selection study. In contrast, Larson and Czerwinski report a frequent backtracking in their Web navigation study. Presumably, the quality of the labels was relatively clear and unambiguous in Miller's menu selection study as compared to those in the Web navigation study. If so, a possible interaction between information structure and label quality might account for the discrepancy between the studies.

To investigate effects of information structure and the quality of the selection labels, we employ a working computational model of Web site navigation. This model simulates a user navigating through a site by executing basic operations such as evaluating links, selecting links, and returning to the previous page. To the extent that the model's behavior is similar to those of human users, we can learn how varying the structure and the quality of labels affect how long it takes to find items in a Web site. By modeling different levels of label ambiguity, our simulations can show the effect of label ambiguity and the extent to which it interacts with the structure of the Web site.

Already computational models have been used to highlight patterns of interactions with a browser (Peck & John, 1992) and report on the accessibility of the site's content (Lynch, Palmiter, & Tilt, 1999). More recent developments include models that predict user behavior or identify usability problems based on the choice of links on each page (Blackmon, Polson, Kitajima, & Lewis, 2002; Chi et al., 2003; Pirolli & Fu, 2003). Constructing and testing a working model complements empirical studies by offering distinct advantages over empirical testing. Empirical studies are generally expensive and time-consuming when they attempt to address the wide range of content, configurations, and user strategies that characterize the Web. In contrast, an implemented model can run thousands of simulated sessions in minutes. Also, empirical studies do not inherently provide explanations for their results and thus make it more difficult to determine how a given result generalizes to other circumstances, whereas a cognitive model embodies and thus describes the underlying processes that produce behavior.

One of our goals is to show the viability of a computational model of information navigation and demonstrate its usefulness in developing a better understanding of how information architecture affects Web navigation. Moreover, we want to use the insight coming from our model for producing sound advice to Web site designers on how to successfully structure their sites. Also, by simulating user actions, including those needed to recover from selecting misleading links, the model estimates navigation costs under a variety of conditions. We ultimately want to use these costs for identifying effective information architectures. With this goal in mind, we call our model MESA (Method for Evaluating Site Architectures).

We start our presentation of MESA by describing how it models Web navigation. Next, we show how MESA's performance explains results from the empirical studies and makes sense of the seemingly contradictory findings. We simulate different levels of link reliability. At one level of link reliability, the MESA's performance is consistent with the Larson and Czerwinski (1998) results. However, for structures with highly reliable links, MESA's performance is consistent with results from menu selection studies.

We then present results from our own user study. We use some of these results to corroborate one of MESA's predictions from the previous section. Finally, we use all of these results to perform a detailed comparison between MESA's performance and the empirical results. We have previously presented descriptions of our model and some initial comparisons to empirical studies (C. S. Miller & Remington, 2000, 2001). For completeness, we fully review the model and describe the initial comparisons before we present the detailed comparisons.

2. MODELING INFORMATION NAVIGATION

Our goal is to simulate common patterns of user interaction with a Web site to provide useful usability comparisons between different site architectures. A model that precisely replicates a user's navigation is not possible, nor do we believe it to be necessary. Rather, a model that employs common usage patterns and simulates them with reasonable time costs can predict and explain benefits of one design over another, such as when it is advantageous to use a two-tiered structure instead of a three-tiered structure.

Because completeness is not possible, process abstraction plays an important role in representing the environment and the human user. Abstraction is used principally to restrict our description of user processing, representing only its functionality. We guide abstraction and model construction with the following principles:

- The limited capacity principle: The model should only perform operations that are within the physical and cognitive limitations of a human user (Broadbent, 1958). For example, limitations of visual attention led us to constrain the model to only focus on (and evaluate) one link phrase at a time (Neisser, 1967). Also, limitations of short-term memory led us to prefer search strategies for our model that require retaining less information over those that require more. In this way we construct our model so that it minimizes memory requirements unless compelling principles or observations indicate otherwise.
- The simplicity principle: The model should make simplifying assumptions whenever possible. The simplicity principle led us to add complexity only if the added complexity was needed to account for observed behavior that is otherwise being systematically misrepresented. For

example the model takes a fixed amount of time to evaluate a link even though the times of human users are certainly variable. Because the model simulates the average user, this simplification will still provide a good approximation given a reasonable estimate of fixed time from human performance data.

- The rationality principle: The model should assume that human cognition is generally rational within the bounds for limited human information processing (Anderson, 1990; Pirolli & Card, 1999). This led us to model options that are the most effective strategy for a given environment unless compelling evidence from human usage suggests otherwise. For example, given the large set of navigation strategies that can operate within reasonable physical and cognitive limitations, we consider the most effective strategies that obey the first two principles.

2.1. Representing a Web Site

MESA interacts with a simplified, abstract representation of a Web browser and a Web site. Each site has one root node (i.e., the top page) consisting of a list of labeled links. Each of these links leads to a separate child page. For a shallow, one-level site, these child pages are terminal pages, one of which contains the target information that the user is seeking. For deeper, multilevel sites, a child page consists of a list of links, each leading to child pages at the next level. The bottom level of all our sites consists exclusively of terminal pages, one of which is the target page. Our examples are balanced trees because we generally compare our results to studies that use balanced tree structures (e.g., Larson & Czerwinski, 1998; D. P. Miller, 1981). However, our representation does not prevent us from running simulations on unbalanced trees, or even on structures involving multiple links to the same page and links back to parent pages.

When navigating through a site, a user must perceive link labels and gauge their relevance to the targeted information. Although the evaluation of a link is a complex and interesting process in itself, we do not model the details of this process. Instead, our interest centers on the consequences of different levels of perceived relevance. As a proxy, we fix a number for each link label, which represents the user's immediately perceived likelihood that the target will be found by pursuing this link. This number ranges in value between 0 (user is certain that selecting it will not lead to the target) and 1(user is certain that selecting it will lead to the target). These relevance values do not necessarily correspond to probabilities because the probability of selecting a link partially depends on which links the user first evaluates on a page. The usage of subjective relevance has its precedence in previous work on exploratory choice (Young, 1998). For work specific to Web navigation, our treatment of link relevance is similar to the concept of residue (Furnas, 1997) or information scent (Pirolli & Card, 1999). It most closely matches the construct of "proximal scent" (Chi, Pirolli, Chen, & Pitkow, 2001).

In an ideal situation, the user knows with certainty which links to select and pursue. Figure 1 represents such a site. The rectangles represent Web pages that contain links (underlined numbers) to child and parent pages. The numbers on links are the link label's relevance to the targeted item, which we define as the user's perceived likelihood that the link is on the path to the target. The top page for this site contains four links where the third link, labeled with a 1.0, eventually leads to the targeted page. Of the eight terminal pages, the page represented by the filled rectangle contains the target information. In our terminology, this example site has a 4 × 2 architecture, where 4 is the number of links at the top level and 2 is the number of links on each child page. For this site, the user need only follow the links labeled with a 1.0 to find the targeted page with no backtracking.

Figure 2 shows an example of a simple two-level site with links whose relevance to the target is less certain. The top page in this figure contains four links labeled with numerical relevance values of .0, .4, .7, and .0 that represent the user's belief that the path associated with a given link contains the target information. As before, a user strategy that merely followed the most likely links would directly lead to the target. Note that the relevance values at any level do not necessarily add to 1. At one extreme, every link at one level could be labeled with a 1, which would represent a user's belief that every link should lead to the desired target.

Figure 3 shows possibly the same site with a different user for whom the meaning of the labels differs from the user in Figure 2. Here the link labels would probably mislead this user away from the target. In this way it is possible to represent sites that differ widely in how well their labels lead users to the targeted item.

2.2. Modeling the Browser and User Actions

To identify common usage patterns important to Web navigation, we use results from a study by Byrne, John, Wehrle, and Crow (1999), who found that selecting a link and pressing the Back button accounted for over 80% of the actions used for going to a new page. Consequently, we have focused on these behaviors and identified component actions underlying them. These actions include

- Selecting a link.
- Pressing the Back button.
- Attending to and identifying a new page.
- Checking a link and evaluating its likelihood.

Figure 1. Site with clear link labels leading to target.

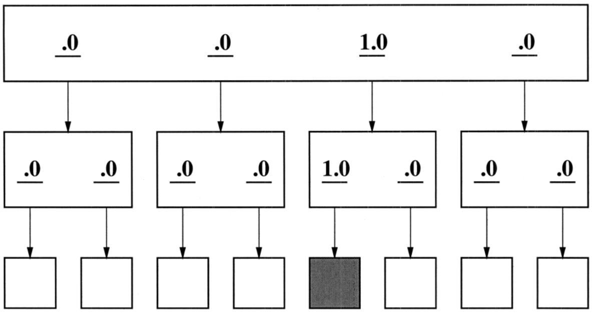


Figure 2. Site with some ambiguity added to link labels.

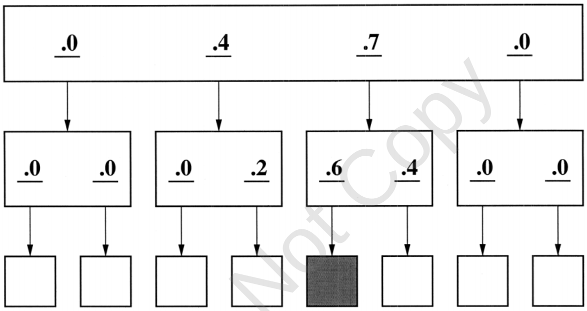
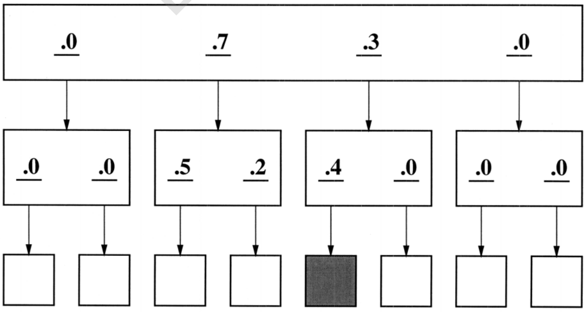


Figure 3. Site with misleading labels. Attending to and identifying a new page.



For present purposes, our model can be further simplified by combining the action of attending to and identifying a new page and folding them into the actions of Selecting a Link and Pressing the Back button because this action only occurs when either of these actions occur. Our revised model has three primitive actions:

- Selecting a link (and attending to and identifying a new page).
- Pressing the Back button (and attending to and identifying a new page).
- Checking a link and evaluating its relevance.

Because of physical and cognitive limitations, only one of these actions can be performed at any one time. Fixed times are assigned to each action to account for its duration during a simulation. The model also simulates changing the color of a link when it is selected so that the modeled user can "perceive" whether the page under this link was previously visited.

2.3. Modeling Navigation Strategies

MESA navigates a Web site by serially executing these three primitive actions. It checks and evaluates links one at a time. Serial evaluation (Neisser, 1967) is motivated by evidence that the human user has a single unique focus of attention (Posner, 1980; Sperling, 1960) that must be directed at the link for this decision (Johnston, McCann, & Remington, 1995; McCann, Folk, & Johnston, 1992).

A user may pursue any number of strategies for evaluating and selecting a link. However, by following the rationality principle, we consider two plausible strategies that minimize the amount of time for finding the target:

- The threshold strategy: The user immediately selects and pursues any link whose probability of success exceeds a threshold.
- The comparison strategy: The user first evaluates a set of links and then selects the most likely of the set.

The threshold strategy is most effective if the first likely link actually leads to the targeted object. The comparison strategy is more effective only if a likely link is followed by an even more likely link that actually leads to the targeted item. Depending on the circumstances, either strategy may be the most effective. However, the comparison strategy requires the user to remember the location and value of the best link to effectively return to it and select it. Consequently, we first examine the threshold strategy on the principle that it requires the fewest computational (cognitive) resources. Only if the threshold strategy provides an insufficient account of user behavior will we consider more complex strategies such as the comparison strategy.

MESA is neutral as to the actual order in which the links are evaluated. The design and layout of a page principally determine which links a user would evaluate first. Any understanding of how page layout and design affect the user's focus could eventually be incorporated into our model. With our current focus on the site structure, MESA's representation establishes a fixed order in which links are evaluated for each run. For our simulations, we can remove the effect of order by randomly ordering links for each run and then taking performance averages across many runs.

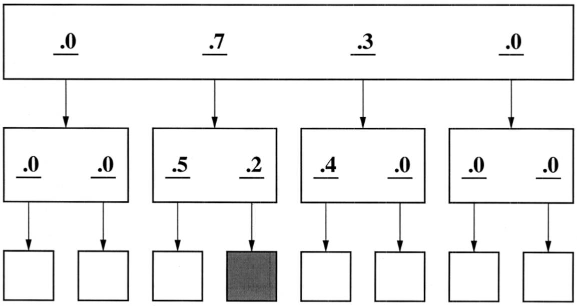
With the appearance of a new page, MESA's threshold strategy first attends to the page, which, if it is a terminal page, includes checking if it contains the target information. If it does not, the model sequentially scans the links on a page selecting any link whose likelihood is equal to or above a fixed threshold (0.5 in the simulations reported here). When a page appears by selecting a link, the process of checking and scanning the page is repeated.

Once MESA detects no unselected links above the threshold value, it returns to the parent page by pressing the Back button and continues scanning links on the parent page starting at the last selected link. It does not scan links it has already evaluated. Determining the last link selected places no demands on memory because the last selected link is easily detected by its color, and many browsers return the user to the location of the last selected link.

So far, for our description, MESA only selects links that will probably lead to the targeted item. However, sometimes the targeted item lies behind ostensibly improbable links and, after some initial failures, human users must start selecting links even if the link labels indicate that they will probably not lead to the targeted item. Earlier versions of our model (C. S. Miller & Remington, 2000) started selecting improbable links only after completing a full traversal of the site. We will call this the traverse-first strategy. However, a more effective strategy would opportunistically select improbable links at a lower tier immediately after trying the more probable links and before returning to a higher tier in the site. We call this the opportunistic strategy (C. S. Miller & Remington, 2001). We adopted this strategy in part based on observed human behavior (see C. S. Miller & Remington, 2001), but also because of its effectiveness.

Figure 4 illustrates how the opportunistic strategy may be more effective. MESA scans across the top page and selects the second link (0.7). On the second level it selects the first link it encounters (0.5). After discovering that this is not the targeted item, it returns to the page on the second level. However, before returning to the top level, it reduces its threshold to 0.1, selects the second link (0.2) and finds the target on the new page. Had the targeted item been elsewhere in the site, the strategy would have MESA back up twice to return to the top level. In order for MESA to restore the threshold to the previous value (0.5), it would need to retain this value across two additional levels of pages. In following our design principle of minimizing memory requirements, we assume that users cannot store and then reset threshold values after traversing multiple pages. Elsewhere we have presented results showing that adding this capability to the model has some marginal impact on three-tiered structures (C. S. Miller & Remington, 2001).

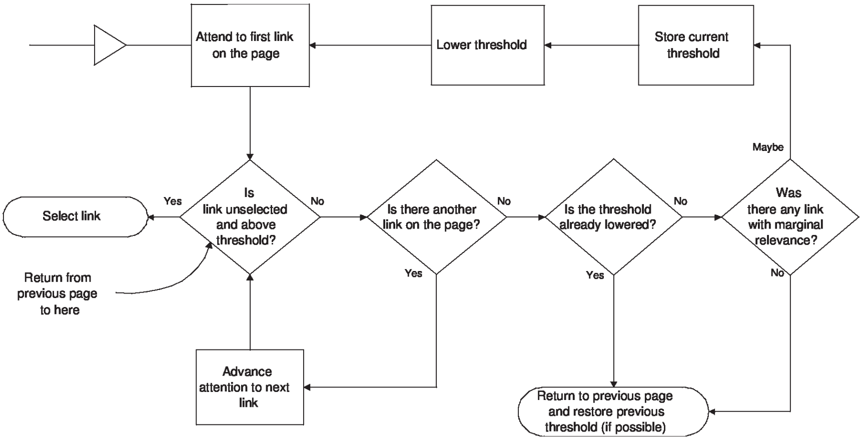
Figure 4. Site for demonstrating the opportunistic strategy.



The opportunistic strategy is a more effective strategy than the traverse-first strategy because it implicitly takes into account the positive evaluation of the parent link, which had indicated that the targeted item was probably under one of the links of the current page. Moreover, the opportunistic strategy explores the less probable links when the cost of doing so is minimal, that is, when the less probable links are immediately available. We further qualify when the opportunistic strategy is used. In some cases, a user may scan a page of links and determine that not even one of these links has the remote possibility of leading to the targeted item (defined as a relevance values of less than 0.1). In this case, our model assumes that the user has the memory to support the realization that rescanning the page would be futile. Instead of employing the opportunistic strategy, the model returns to the parent page. This memory of knowing that the page has nothing worthwhile only lasts as long as the model remains on the current page. Thus, if MESA leaves the page and then returns to this same page, the model must assume that the page may be worth rescanning and the opportunistic strategy is employed. This qualification is also consistent with our design principles in that it contributes to an effective strategy while minimizing memory resources.

Figure 5 provides a flowchart for the major actions and decisions of the opportunistic strategy. The flowchart starts when MESA attends to a new page of links and leaves the flowchart by selecting a link or returning to the previous page. Starting at the first link on a page, MESA iteratively evaluates each link. If the link relevance exceeds the current threshold, it selects that link and the process starts again at the new page. When MESA reaches the last link on the page, the flowchart shows how MESA may rescan the page at a lower threshold unless its memory indicates that it did not pass any marginally relevant links (note that this memory is lost if it leaves the page, in which case it will always rescan if it can lower the threshold). When MESA returns from a page, it continues the scan starting at the last selected link.

Figure 5. Flowchart summarizing the opportunistic strategy.

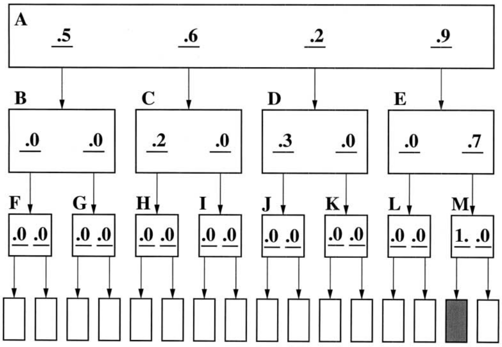

Attend to first link on the page
Lower threshold
Store current threshold
Maybe
Select link
Yes
Is link unselected and above threshold?
No
Is there another link on the page?
No
Is the threshold already lowered?
No
Was there any link with marginal relevance?
Return from previous page to here
Yes
Yes
No
Advance attention to next link
Return to previous page and restore previous threshold (if possible)

2.4. Detailed Example

To illustrate how MESA navigates a structure, we review a detailed example. Figure 6 depicts a structure that we have deliberately created to demonstrate a complete set of the model's properties. The bottom level of pages consists of only reliable link labels, which corresponds to most structures we will study in this article. To keep the example simple, we reduced the number of links per page to be less than what is in a typical Web site. This example has some deceptively relevant links and MESA needs to perform extensive backtracking before it finds its target. The top level has two deceptively relevant links (valued at .5 and .6). The first of these requires minimal backtracking, but the second link leads to a page with one marginally relevant link (valued at .2). Here we will see that MESA undergoes a substantial amount of backtracking before it returns to the top and selects the link that eventually leads it to the target.

Figure 7 presents the component actions of the model as it navigates this structure. For this trace, we will assume that the original threshold is set to .5 and the secondary threshold is set to .1. At the top level, it first selects the deceptive link valued at .5. When Page B appears, it first scans the page at the original threshold. Because no links are above the threshold, none are selected. Moreover, because no marginally relevant links were encountered, MESA backs up to Page A without rescanning the links. Because MESA only descended one page and did not lower the threshold, it returns to Page A with the original threshold, valued at .5.

Figure 6. Structure for detailed example.



MESA then evaluates the deceptively relevant link valued at .6. After Page C appears, it first scans the page at the threshold of .5. On the first scan, no links are above the threshold, but it does note the marginally relevant link valued at .2. Finding marginally relevant links on the first scan, it lowers the threshold to .1 and scans Page C again. This time it selects the link valued at .2, which leads to Page H. Page H requires only one scan to determine that it has no relevant links. MESA returns to Page C. It evaluates the second link valued at .0 before returning to Page A.

By the time MESA returns to Page A, it has visited two pages across two levels. With its memory limit, it can no longer restore the threshold to its previous value. It thus continues the scan of Page A at the lower threshold value of .1. It now selects the link valued at .2, which leads to Page D and Page J before it backtracks to Page A. Finally, the next link, valued at .9, leads to the target.

This navigation requires 19 link evaluations, 8 link selections, and 5 Back actions. In the next section, we provide time constants to each of these actions to predict the total navigation time. In addition to the user actions, both the link selection and the Back action should include the system response time needed for having the next (or previous) page appear. The model can thus account for a slow network response by using larger time values for the action of link selection.

This example also illustrates how structure, link relevance, and cognitive limitations can interact to increase the number of actions needed to find a target. Compared to a two-tiered structure, a three-tiered structure has fewer links per page. When the link labels are reliable, the three-tiered structure may provide an efficient path to the target. However, this three-tiered example has deceptively relevant links at the top level. In the first case (top link valued at .5) produces minimal backtracking, but the second deceptively relevant link (valued at .6) leads to a page with marginally relevant links that require a second scan, an additional page selection, and backing up two lev-

Figure 7. Simulation trace.

Action	Page	Comment
Eval .5	A	Link is deceptively relevant
Select .5	A	Link is at threshold of .5
Eval .0	B	Link is not above threshold
Eval .0	B	Link is not above threshold
Back to A	B	Threshold stays at .5
Eval .6	A	Link is deceptively relevant
Select .6	A	Link is above threshold of .5
Eval .2	C	Link is marginally deceptive but below threshold
Eval .0	C	Too low, lower threshold and rescan page
Eval .2	C	Rescanning page with lower threshold of .1
Select .2	C	Link is above lower threshold
Eval .0	H	First link on H
Eval .0	H	Second link on H
Back to C	H	No relevant links, no rescan
Eval .0	C	Check if marginally relevant
Back to A	C	Can no longer recall previous threshold
Eval .2	A	Link after last selected link
Select .2	A	Link is above .1 threshold
Eval .3	D	
Select .3	D	
Eval .0	J	First link on J
Eval .0	J	Second link on J
Back to D	J	No relevant links, no rescan
Eval .0	D	
Back to A	D	At low threshold, no rescan
Eval .9	A	Link after last selected link
Select .9	A	Link is above .1 threshold
Eval .0	E	
Eval .7	E	
Select .7	E	
Eval 1.0	M	
Select 1.0	M	Arrive at target

Note. Summary of actions: 19 link evaluations, 8 link selections, 5 Back actions.

els. Cognitive limitations contribute to additional navigation costs in three ways. First, only one link can be evaluated at a time, which causes the model to evaluate and select deceptively relevant links before it evaluates the highly relevant link that leads to the target. Second, Page C needs to be scanned a second time to find the marginally relevant link. Third, because of a memory limitation, the selection criterion at the top page is lost after traversing multiple levels. The lower criterion causes the model to select marginally relevant links before all highly relevant links are evaluated.

2.5. Relation to Other Approaches

Our example shows how the relevance of link labels plays an integral role in how MESA predicts user behavior. Some other approaches do not explicitly model the relevance of link labels (Bernard, 2002; Lynch et al., 1999). In these cases, the quality or distribution of label relevance in a site is not a factor in the model's predictions. These idealized models could make valid relative predictions for sites with less-than-ideal labels if any degradation in label quality equally affected all structures. However, we suspect that there are important cases where structures are not equally affected. First of all, changing the structure may force the designer to remove helpful links or add misleading links. In this way, an otherwise ideal structure may perform poorly if its structure does not fit a good choice of selection categories. Secondly, even if the compared structures supported the same level of label quality, we believe that some structures would be more affected by having less reliable labels for its selections. We will further explore this second point in the empirical sections of our presentation.

Other predictive models of information navigation do incorporate label relevance in their processes or calculations. For the Cognitive Walkthrough for the Web (CWW), link relevance is the principal consideration for making predictions on the accessibility of targeted items (Blackmon et al., 2002). CWW uses Latent Semantic Analysis (LSA) as an automated method for assessing link relevance. In the next section, we further discuss LSA and other methods for assessing link relevance. Using LSA, CWW identifies unfamiliar, confusable, and competing links to identify potential navigation problems. To the extent to which link relevance is the dominant contributor to a structure's accessibility, CWW provides a useful method for selecting and evaluating structures. However, the structure of the site may be an important factor in determining the cost of selecting the wrong link. CWW does not account for this cost. In contrast, MESA explicitly calculates the cost of selecting a misleading link by simulating the actions needed to recover from the mistake. We will also explore this cost in the empirical sections of our presentation.

The Bloodhound Project (Chi et al., 2001; Chi et al., 2003) explicitly models label relevance and the abstract structure of a site. It uses a spreading activation model to simulate user navigation, where the level of activation on a page depends on the proximal scent of the links leading to the page. The cost of backtracking is considered by employing the "Information Scent Absorption Rate" method (Chi et al., 2003), which returns simulated navigation back to the starting page after exhausting a dead-end.

Unlike Bloodhound, MESA's navigation strategies are additionally constrained by some cognitive limitations. In our detailed example, we saw how cognitive limitations may incur additional costs when backtracking occurs among multiple levels. We next see how these limitations play a role in predicting navigation times across different structures.

2.6. Simulation Parameters

Time Constants

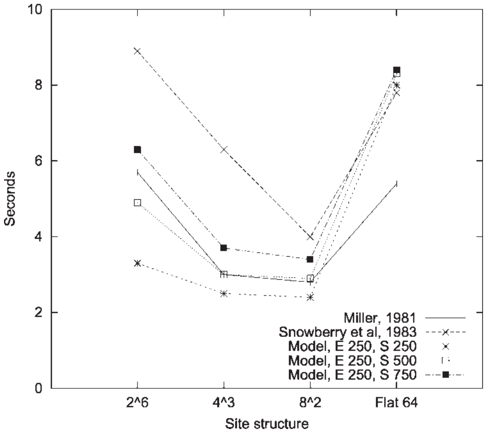
Because we are interested in having our model estimate navigation times for finding an item, we need plausible time constants for each of the component actions (i.e., link evaluation, link selection, and pressing the Back button). In previous work, we established plausible estimates for link evaluation and link selection (C. S. Miller & Remington, 2000). We derived these constants by comparing MESA to results from hierarchical menu selection studies. D. P. Miller (1981) reported that humans searching through an 8 × 8 menu system took slightly less than 3 sec on average. Because selection errors occurred less than 1% of the time, we conclude that the system's selection labels were accurate, clear, and unambiguous. We simulated these results by creating a model of an 8 × 8 site with clear, unambiguous labels. Nodes that led to the target were given a relevance value of 1, others were given a relevance value of 0. Because no backtracking occurs, there are only two timing costs that need to be established: evaluating a link label and selecting a link (recall that selecting a link includes the time to display and to start examining the next page).

Using ranges of costs from 250 msec to 750 msec at increments of 250 ms, we ran simulations to find out which combinations produced a good match to D. P. Miller's result of slightly less than 3 sec. The cost settings closest to the Miller result were 250 msec for link evaluation and 500 msec for link select, which produced a total search time of 2.9 sec. Settings of 250/250 and 250/750 produced total search times of 2.4 sec and 3.4 sec, respectively, which are also close to the 8 × 8 total time reported by Miller. Other cost settings were significantly further away from the Miller result.

We then took the three best-fitting cost settings on the 8 × 8 (8^2) comparison and evaluated their performance on other structures tested by D. P. Miller. These alternate structures include 2 × 2 × 2 × 2 × 2 × 2 (2^6), 4 × 4 × 4 (4^3), and a flat structure of 64 selections. We also compared our results to those presented by Snowberry, Parkinson, and Sisson (1983), who also ran studies on menu systems with these same structures.

The results of our comparisons are shown in Figure 8. The results for D. P. Miller, Snowberry et al. (1983), and the three sets of simulations all show the same qualitative pattern, namely that the 8 × 8 structure produces the fastest search times compared to the other structures. As for absolute times, the parameter values of 250/500 model matched the Miller data the closest and will serve as the initial estimates for our simulations. Because pressing the Back button is an operation comparable to making a link selection, we initially use the same time constant as for link selection.

Figure 8. Comparison between menu selection results and simulations.



The use of time constants for predicting human interaction times is well established (Card, Moran, & Newell, 1983; Lee & MacGregor, 1985). Our initial estimates are probably lower bounds on the range of plausible time constants. An average link evaluation time of 250 msec assumes that a user would have to process the whole link label with each saccade, which has been estimated to last 230 msec on average (Card et al., 1983). The estimate of 500 msec for selecting a link and attending to the next page assumes that the user already has the pointing device in position and that the system response time is negligible.

Setting Relevance Values for Link Labels

The comparison to menu selection results assumes ideal links. That is, the model need only follow a "1" to successfully find its target page without any backtracking. Although this assumption may be appropriate for simulating menu selection studies where little backtracking occurred, it does not model situations, which include many Web sites, where users frequently select the wrong links and need to backtrack.

Our method for modeling less reliable link labels is to start with a structure consisting of clear labels. We then perturb the values of the ideal links with the use of a random variable from a standard normal (Gaussian) distribution (M = 0, SD = 1). In particular, we change all link values of zero to the following:

$$|g| * n$$

All link values of one are changed to the following:

$$1 - |g| * n$$

In these formulas, g is produced from a standard normal distribution, which is commonly available as a library routine in programming environments. To achieve the distance from the ideal value, the absolute value of g is multiplied by n, the noise factor multiplier (equivalent to increasing the variance of the normal distribution). Occasionally this formula produces a value outside the range from zero to one. In these cases, the formula is iteratively applied until a value within the range is produced.

The noise factor n models the level of label reliability in the site. By increasing the value of n, we increase the probability that the model will skip a link that leads to the target and also increase the probability that it will select a link that does not lead to the target. For example, when n equals .3, a label leading to a target has a 90.4% chance of being assigned a relevance value greater than .5. If we establish a selection threshold of .5, a link leading to the target will have a 90.4% chance of being selected. A link that does not lead to the target (a foil link) has a 9.6% chance of being selected at this threshold.

Figure 9 shows the probabilities that an evaluated link will be selected under a variety of selection thresholds and noise factors. For example, at a noise level of .3, if the threshold has been reduced to .1 and the model is evaluating a foil link (i.e., a link that does not lead to the target), there is a 74% chance it will be selected. For our first set of simulations, we will use a primary threshold of .5 and a secondary threshold of .1.

When considering a selection probability, it is important not to confuse it for a probability that the model will move from the current page to the linked page. This page transition probability is not easily calculated and depends on the order of the links, the relevance values of other links, and the current threshold. Moreover, as we have seen in the example trace in Figure 7, the current threshold may depend on what other pages have already been visited. For this reason, a simple Markov model that has its states correspond to pages would not be able to fully account for the model's behavior.

Another way of setting relevance values for link labels involves the use of human raters. Given the name of an item and a link label in an actual Web site, a person can provide a rating of how likely he or she would expect to find this item by selecting the link. In one case, the human raters can be expert judges who try to anticipate how users will interpret the link labels. Alternatively, the human raters may be recruited from the user population and asked about their probable actions when viewing linked labels. Later in this article, when we perform detailed comparisons between human results and those from the model, we will use our own assessments to estimate link relevance.

A final method for setting relevance values could make use of automated methods that estimate the semantic distance between two phrases, that is, the name of the item and the link label. LSA is one method that provides such a distance (Landauer & Dumais, 1997). This metric is derived from how fre-

Figure 9. Selection probabilities by threshold and noise level.

		Selection Probabilities	
Threshold	Noise Factor	TargetLink	FoilLink
Primary (0.5)	.1	1.00	.00
	.2	.99	.01
	.3	.90	.10
	.4	.79	.21
	.5	.70	.32
Secondary (0.1)	.1	1.00	.32
	.2	1.00	.62
	.3	1.00	.74
	.4	.98	.81
	.5	.95	.85

quently words (and words discovered to be related to these words) co-occur in targeted texts. As a measure of relatedness, this metric can be interpreted as the likelihood that the item will be found if the label's link is selected.

There has been some exploration of how LSA can be used for automatically measuring the quality of labels in a user interface (Soto, 1999). More recently, LSA has been applied to identify potential usability problems in Web sites (Blackmon et al., 2002). One current limitation of LSA is that the user target needs to be specified as 100 to 200 words of text to produce accurate predictions (Blackmon, Kitajima, & Polson, 2003). Other approaches have also used distance measures based on word co-occurrences in text documents (Pirolli & Card, 1999) or the World Wide Web itself (Pirolli & Fu, 2003). To be effective, the content of the text documents needs to correspond to the conceptual knowledge of the users.

3. SIMULATIONS

In this section, we first explore the effect of label ambiguity on the structures used in the Larson and Czerwinski (1998) study. We see that MESA produces behavior that is consistent with their results once we apply a sufficient amount of noise to the link values. We then perform more detailed comparisons between results we collected ourselves and MESA.

3.1. Modeling Structure and Link Ambiguity

Using our model MESA, we conducted simulations using the threshold strategy for link selection with the opportunistic strategy for backtracking. Sites were constructed by randomly placing the target item at one of the terminal pages and assigning a value of 1 (links leading to the targeted item) or 0 (all other links). Link values were then perturbed by Gaussian noise as described earlier. The noise was not applied to the bottom level, which leads to the terminal pages. Although not necessarily plausible for all Web sites, this treatment corresponds to the sites used by Larson and Czerwinski (1998) because their participants could clearly tell whether the link's label matched the text of the targeted item. Figure 10 shows a sample 4 × 2 × 2 architecture generated with a noise factor of .3.

For each site architecture (8 × 8 × 8, 16 × 32, and 32 × 16), 10,000 simulations were run using the following time costs: 250 msec for evaluating a link, 500 msec for selecting a link, and 500 msec for returning to the previous page (pressing the Back button). Following Larson and Czerwinski (1998), any run lasting more than 300 sec was coded as lasting 300 sec.

Figure 11 shows the calculated mean times of the simulation runs. The simulated results are displayed with connected lines. Not surprisingly, the time needed to find a target increased with link ambiguity. What is more interesting is how link ambiguity interacts with site structure. The 8 × 8 × 8 architecture produced slightly faster times at low levels of noise but substantially slower times at noise levels above .2. At these higher noise levels the results are consistent with the human users (which are indicated with arrows in the figure). At noise levels of .4 and higher, simulated times were faster with the 16 × 32 architecture than the 32 × 16 architecture. This difference was also noted in the study with human users, albeit not reported as statistically significant.

At a noise level of .4, the simulation results closely match the human results in absolute terms: 62 sec (compare to 58 sec for humans) for 8 × 8 × 8, 43 sec (compare to 46 sec) for 32 × 16, and 35 sec (compare to 36 sec) for 16 × 32. It appears that the .4 serves a good parameter estimate describing the amount of label ambiguity in the sites used by Larson and Czerwinski (1998).

3.2. Impact of Time Costs

Although changing the time costs (250 msec for link evaluations and 500 msec for link selection and returning to the previous page) will affect absolute simulation times, it is less clear if different time costs will change which architecture produces the fastest times. For example, one may wonder if the 8 × 8 × 8 architectures would still produce the slowest times if the link selection cost were doubled, which may be the case for a slower Internet connection.

To explore the impact of time costs, we looked at the number of link evaluations, link selections, and page returns. If independent counts of these actions correlate with the aggregate simulation time, we conclude that varying the time costs have minimal impact on the relative performance of the different architectures. For example, if the 8 × 8 × 8 requires more evaluations,

Figure 10. Site with no label noise on the bottom level.

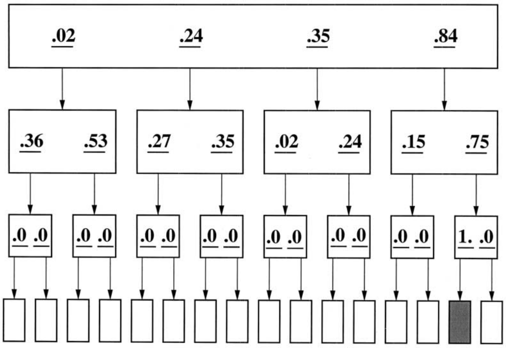
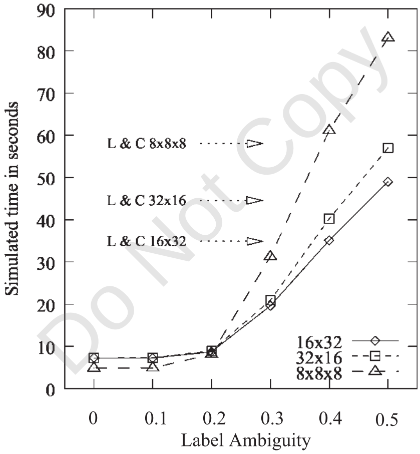


Figure 11. Performance as a function of link ambiguity and architecture.



more selections, and more returns than the other architectures, we know that 8 × 8 × 8 will produce slower search times regardless of the time costs.

Looking at the number of evaluations, selections, and returns, the 8 × 8 × 8 architecture required more of each action (173, 17, and 19, respectively) at the .4 noise level than the 16 × 32 (125, 3, and 5) and the 32 × 16 (134, 6, and 8). Further analysis revealed that this relationship holds across all but the lowest noise levels (.2 and less). We conclude that changing the time costs, at least for these structures, has no effect on the relative comparisons provided that the noise level is at least .3. More generally, it suggests that increasing label ambiguity equally increases the number of all three actions. This conclusion seems reasonable because there needs to be corresponding link evaluations and returns for each incorrectly selected link.

3.3. Impact of Bottom-Level Noise

These results were from simulations where the bottom level of links have unambiguous labels. Although this corresponds to the sites constructed for the Larson and Czerwinski (1998) study, this assumption does not hold for many real Web sites. In particular, people often do not search for a specific item, but need to visit the target page before realizing they have found what they are looking for. To explore the consequences of having ambiguous links at the bottom level, we ran simulations where the noise-producing process was applied to every label in the site.

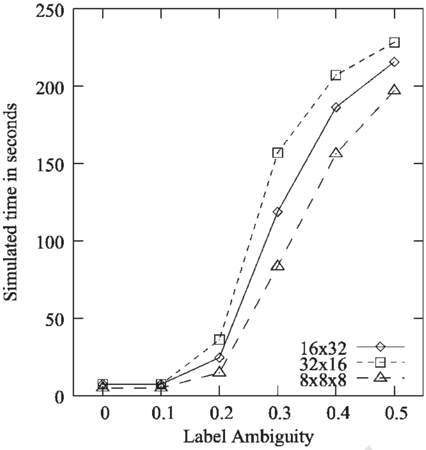
Figure 12 shows the results. Not surprisingly, the addition of noise on the bottom level increased the absolute times for all structures (note that a larger range for the y-axis is required to plot the results). More important, the three-tiered 8 × 8 × 8 structure this time produced the fastest results for all noise levels. With the addition of ambiguity at the bottom level, the model's behavior is now consistent with empirical and theoretical findings for menu selection studies, which indicate that pages with 8 choices per page are better than those with 16 or more choices per page. The model's behavior further suggests that the Larson and Czerwinski result does not generalize to structures where label ambiguity is equally distributed throughout the structure.

3.4. Discussion

Using structures with clear labels that reliably lead to the target, our simulations found the target faster in the three-tiered structure than in the two-tiered structures. This simulated behavior is consistent with the menu selection studies. It is also consistent with the theoretical analysis provided by Lee and MacGregor (1985). Like MESA, their analysis assumes a linear self-terminating search. Using reasonable values for the time needed for evaluating each label and for selecting and loading each menu, they determined an optimal range of 4 to 13 selections per menu. The 8 links per page of the three-tiered structures fall within this range, whereas the 16 and 32 links per page of the two-tiered structures do not.

Of course, for a sufficiently large selection cost (which includes the time to load the page), the optimal range of links per page would increase and then favor the two-tiered structures. Using the assumptions of Lee and MacGregor, we can determine at what point the two-tiered structures would become more efficient. For the average case, the three-tiered structure would require 4.5 evaluations per level for a total of 13.5 evaluations. For the two-tiered structures, 8.5 and 16.5 evaluations are needed at their two levels for a total of 25 evaluations. The three-tiered structure requires three selections and the two-tiered structures each require two selections. If m is the ratio of selection cost to evaluation

Figure 12. Simulated performance with noise at all levels.



cost, the following equation reveals when the three-tiered structures and the two-tiered structures would yield equivalent navigation times:

$$13.5 + 3m = 25 + 2m$$

$$m = 11.5$$

This calculation indicates that the cost of selecting and loading a page would need to be greater than 11.5 times the cost of evaluating a label to give the advantage to the two-tiered structures. Assuming a very fast evaluation time of 250 msec per label, the selection and loading cost would need to be substantially greater than 2.9 sec. This is plausible for slower network connections. However, this analysis assumes that the amount of time needed to load a page is a constant. For slower network connections, it is likely that load times may vary, increasing with the number of links per page, and thus penalize the two-tiered structures further. If so, it is not clear whether these two-tiered structures would be optimal under any plausible timing assumptions, that is, for when the structures' labels reliably lead users to the target with the minimal number of link selections.

However, both the Larson and Czerwinski (1998) results and our simulation results suggest that the theoretically optimal number of links does not apply to structures whose labels are sufficiently ambiguous or unreliable at the top level(s) but clear at the bottom level. In these cases, the two-tiered structures produced faster navigation times than the three-tiered structure. The underlying behavior of our model offers a possible explanation. As we noted when presenting the detailed example in Figure 6 and Figure 7, an incorrect selection at the top level followed by an incorrect selection at the middle level incurs an additional cost of double-checking the other middle-level links after returning from the third level. This additional cost does not occur for the two-tiered structures, provided that the links at their secondary levels are sufficiently clear so as not to cause any selection errors.

4. COLLECTING DETAILED HUMAN PERFORMANCE RESULTS

Our simulations suggest that there is an interaction between structure and label ambiguity, at least when label ambiguity is varied at all levels but the bottom level. In particular, the simulations predict faster search times for the three-tiered 8 × 8 × 8 structure when category labels are clear, but faster search times for two-tiered structures (i.e., 16 × 32 and 32 × 16) when labels are ambiguous. To our knowledge, there are no previous empirical studies that explore possible interactions between structure and label ambiguity.

In this section, we present results from our own empirical study, where we purposely selected targets that lay behind categorical labels of varying reliability. Because we use the actual names of the targets at the bottom level, we will further test our model's predictions for when there is no ambiguity at this level. With the results of this study, we are able to further explore the interaction between structure and label ambiguity and perform detailed comparisons between the model's performance and that of human participants. A preliminary analysis of these results was previously presented in C. S. Miller and Remington (2002).

For this study, we used a three-tiered structure that closely approximates the 8 × 8 × 8 structure used by Larson and Czerwinski (1998) and our simulations in the last section. From this structure, we derived two two-tiered structures, one of which closely approximates the 32 × 16 structure. To test our predictions, we focus on the two structures that best correspond to the Larson and Czerwinski study, but in the following section we use all of the results for further evaluating the model.

4.1. Method

Participants

Forty-five participants were recruited from class announcements and student e-mail lists at DePaul University. The classes and e-mail lists only included students who were at least in their second year of study. The call for participation required at least 10 hr of personal usage on the Web and an age of at least 18 years. As students at DePaul University, these participants had frequently used the Web to look up schedule information and register for courses.

Materials

The Web sites were constructed using items and categories found in a discount department store. Of the categories, there were 6 high-level categories and 37 low-level categories. Examples of items are a tripod grill, a butane lighter, and a hand mixer. Examples of the 6 high-level categories are sporting goods and hardware. Examples of the 37 low-level categories are camping accessories and kitchen gadgets. A Web server dynamically constructed a site hierarchy from these categories and items. The three-tiered structure was created from categories at both levels, where the top-level page had 6 links, the pages at the second level had an average of 6.17 links, and the pages at the bottom level had an average of 13 links, each leading to the items. Two-tiered structures were created by either omitting the top-level categories or the bottom-level categories. Omitting the top-level categories produces a two-tiered 37 × 13 structure, which has 37 links at the top-level and an average of 13 links per page at the bottom level. Similarly, omitting the bottom-level categories produces a two-tiered 6 × 80.8 structure. The entire structure and its labels are presented in the Appendix.

Procedure

Using a between-groups design, each participant was randomly assigned to search in one of the three structures. Regardless of structure, each participant was asked to look for the same eight items. We chose target items based on our subjective assessments and those of a third judge who was knowledgeable of the study. We predetermined that two of these are clearly categorized at both levels and that two of these items are ambiguously categorized at both levels. The remaining four items were judged to have ambiguous labels at one level but not the other level.

The Web server randomized the order of search targets for each participant and created a new Web site for each search by randomizing the order of the links on all of its pages. Every time a participant requested a new page by selecting a link, the Web server automatically recorded the name of the selected link and the time the link was selected. If the participant took longer than 4 min, the server asked the participant to look for the next targeted item.

4.2. Results

Figure 13 shows the summary of human performance on all eight tasks for each of the three structures. Each mean in this table pools data from 15 participants. For cases when the target was not found, the search time was recoded as 4 min. For link selections, three selections were the minimum for the

Figure 13. Summary of human performance on navigation tasks.

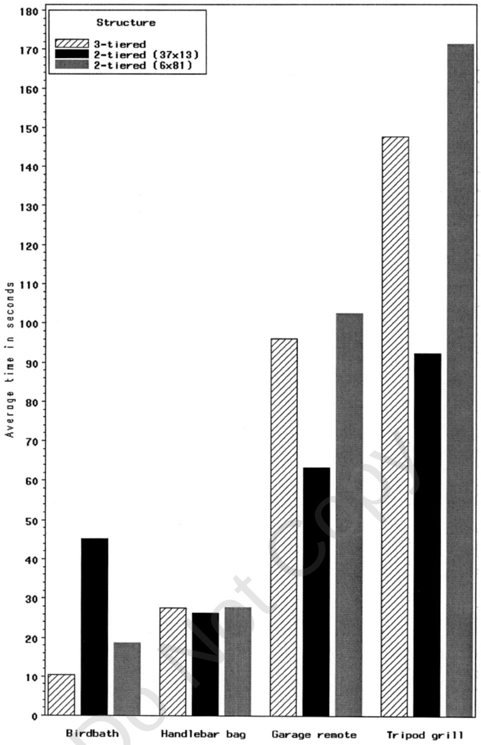
			Time (Seconds)		Selections	
Label Reliability	Target name	Approximate Structure	M	SD	M	SD
Links judged reliable at both levels	Birdbath	6 × 6 × 13	10.3	3.9	3.1	0.4
	Birdbath	37 × 13	45.1	57.4	3.3	3.1
	Birdbath	6 × 81	18.6	32.9	2.4	1.5
	Handlebar bag	6 × 6 × 13	27.6	32.1	4.5	2.5
	Handlebar bag	37 × 13	26.2	28.7	2.7	1.4
	Handlebar bag	6 × 81	27.8	21.0	2.1	0.3
Links judged reliable at neither level	Garage remote	6 × 6 × 13	96.0	69.4	13.7	8.1
	Garage remote	37 × 13	63.5	26.6	4.3	1.2
	Garage remote	6 × 81	102.6	75.5	5.1	3.4
	Tripod grill	6 × 6 × 13	147.8	77.0	19.2	12.6
	Tripod grill	37 × 13	92.4	52.3	7.5	3.3
	Tripod grill	6 × 81	171.3	61.2	8.0	4.9
Links judged reliable at top level	Chopsticks	6 × 6 × 13	49.9	61.7	5.7	4.8
	Chopsticks	37 × 13	75.9	65.6	5.4	4.7
	Chopsticks	6 × 81	32.3	46.5	2.4	1.5
	Hand mixer	6 × 6 × 13	68.3	64.7	8.1	5.6
	Hand mixer	37 × 13	40.5	30.2	3.7	1.8
	Hand mixer	6 × 81	109.7	62.9	5.1	2.7
Links judged reliable at second level	Shower organizer	6 × 6 × 13	75.4	53.6	8.0	4.4
	Shower organizer	37 × 13	19.1	10.3	2.1	0.3
	Shower organizer	6 × 81	75.2	52.5	3.8	1.7
	Tire scrubber	6 × 6 × 13	21.2	13.1	3.7	0.9
	Tire scrubber	37 × 13	32.4	40.6	3.4	2.6
	Tire scrubber	6 × 81	33.0	42.9	3.6	3.6

three-tiered structure whereas the two-tiered structures only require a minimum of two selections.

The complete set of results will be compared to simulated results in the next section. Here we focus on tasks pertinent to the predictions in the previous section. They include the two targets (i.e., birdbath and handlebar bag) whose labels were previously judged unambiguous at both categorical levels and the two targets (i.e., garage remote and tripod grill) whose labels were judged ambiguous at both levels. Graphed comparisons of average times in seconds are shown in Figure 14. The times across all three structures were fastest for the unambiguous targets (the birdbath and the handlebar bag) and slowest for the ambiguous targets (the tripod grill and the garage remote).

For comparisons, we focus on results that address the predictions from the simulations by considering the structures that approximate the structures from our simulations. These are the three-tiered structure (its 6 × 6.17 × 13

Figure 14. Human performance by target and structure.



structure approximates an 8 × 8 × 8 structure) and the two-tiered structure with bottom-level categories (its 37 × 13 structure approximates a 32 × 16 structure). Because the variances between our groups were often significantly different, we employed a Satterthwaite, separate variances, t test to analyze our comparisons.

The birdbath was found significantly faster in the three-tiered structure (M = 10.3, SD = 3.9) than in the two-tiered structure (M = 45.1, SD = 57.4), t(14.1) = –2.34, p = .035, two-tailed. In contrast, the tripod grill was found significantly faster in the two-tiered structure (M = 92.4, SD = 52.3) than in the three-tiered structure (M = 147.8, SD = 77.0), t(24.7) = 2.30, p = .030, two-tailed.

The difference for the garage remote was less reliable, t(18) = 1.70, p = .107, two-tailed, and there was no significant difference for the handlebar bag, t(27.7) = .12, p = .904, two-tailed.

Although the birdbath and the handlebar bag were prejudged to be unambiguously categorized targets, not all participants took the shortest route. For example, many participants first looked for the handlebar bag under Hardware before choosing the correct category, Sporting Goods. For these participants, the handlebar bag lies behind ambiguous labels and does not appear to match our assessment as an unambiguously categorized target.

We consider an alternate method for selecting tasks that better corresponds to individual assessments. Instead of relying on judged assessments, the quality of the labels could be measured by counting the number of link selections a participant took to find the item. Tasks with clear labels could be identified as those for which participants only performed a minimal number of link selections. Likewise, tasks with ambiguous labels could be identified as those requiring the largest number of link selections.

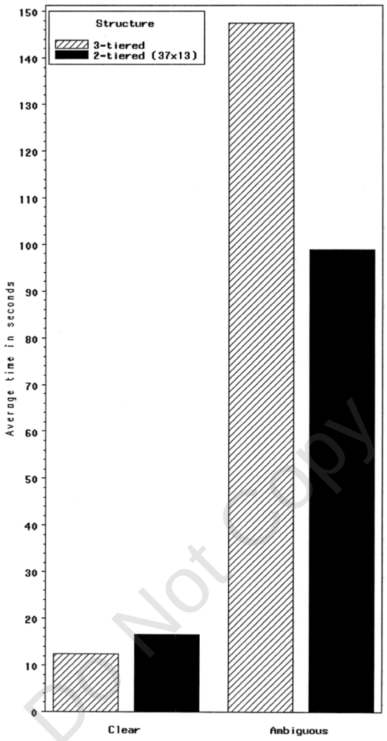
We performed a second analysis using this measure of label reliability. For each participant, we ranked the tasks by the number of link selections needed to find the target. We identified "clear label" tasks as those with the first and second fewest number of selections. For tasks with tied ranks, we averaged their navigation times before creating the "clear label" task average for each participant. With this method, at least two of the eight tasks were included in the analysis for each participant. Similarly, we created an "ambiguous label" task average for each participant using tasks with the most link selections.

Figure 15 shows the results where link ambiguity is defined by the relative number of link selections. We used a pooled variance t test for the analysis. For tasks requiring the fewest link selections, targeted items were found faster in the three-tiered structure (M = 12.42, SD = 3.75) than in the two-tiered structure (M = 16.56, SD = 4.67), t(28) = –2.67, p = .012, two-tailed. For tasks that took the most link selections, targeted items were found faster in the two-tiered structure (M = 98.9, SD = 45.0) than in the three-tiered structured (M = 147.4, SD = 45.7), t(28) = 2.93, p = .007, two-tailed.

4.3. Discussion

The comparisons for which there was a significant difference (i.e., $p < .05$) were all consistent with the theoretical prediction, namely that items whose link labels are unambiguous are generally found faster in a three-tiered structure (approximately 8 links per page) than in a two-tiered structure (approximately 32 links per page at the top level and 16 links per page at the bottom level) and that items whose link labels are ambiguous are generally found more slowly in the three-tiered structure than in the two-tiered structure. In regard to the results by Larson and Czerwinski (1998), their finding that 32 × 16 structures produce faster times than 8 × 8 × 8 structures seems to generalize to similarly sized structures provided that the targeted items are not clearly classified from the perspective of their users.

Figure 15. Human performance with label ambiguity defined by the number of link selections.



The information structures were derived from the actual organization of a department store. As is the case with most structures, some pages contain more items than other pages. In this sense, the structures are more realistic than the evenly balanced structures used by Larson and Czerwinski (1998). One consequence of starting with an unbalanced structure is that the derived two-tiered structure may be unduly affected by the number of items per page. For example, to find the birdbath in the three-tiered structure, the human user first selects Garden among 6 items, then Patio Accessories among 3 items, and finally birdbath among 11 items. For the comparable two-tiered structure, the user selects Patio Accessories among 37 items before selecting birdbath among 11 items. In this case, the three-tiered structure has a potential time advantage because its 6 × 3 structure indexes potentially fewer items than the corresponding 37 items in the two-tiered structure. On the other hand, the handlebar bag favors the two-tiered structure (37 × 10) over the three-tiered structure (6 × 7 × 10) because 37 categories indexes fewer items than 6 × 7 categories.

The discrepancies caused by the unbalanced structures are arguably not large enough to change the direction of the significant differences. For the birdbath, adding four more links to the page of three links would not reverse the advantage of the three-tiered structure. If we assume that on average two of these four links will be evaluated before the correct link is selected, we can calculate the additional time by multiplying 2 by the link evaluation time, which we have set to 250 msec for our model. The resulting time of 500 msec is an order of magnitude less than the observed difference in times between these two structures.

The difficulty of fairly comparing the effect of different structures highlights the advantage of a computational model. With a model, the experimenter can expressly set ambiguity factors and site structure to rule out any confounding factors. Of course, this demands a model that has been adequately validated against human data. In the next section, we use the detailed results from our empirical study to further validate the model.

5. DETAILED SIMULATIONS AND COMPARISONS

Our empirical study involved 45 human participants with each of them navigating a site looking for a total of eight items in one of three structures. If we average the results across the 15 participants for each target in each structure, we produce 24 mean times (as presented in Figure 13). In this section, we will compare these times with those predicted by MESA. Unlike our previous simulations, MESA runs on a site representation that has a direct correspondence to the site that each participant navigated during their search tasks. Because we saved each site structure, search task and label orderings for each of the 360 (15 × 8 × 3) tasks, we are able to present the model with the same site representations.

For determining numerical values that represent label relevance, we use the judged ratings that had been collected to choose the targets in the user study. Relying on ratings from only three judges represents a cost-effective method for quickly estimating the relevance of each label with respect to each target.

5.1. Simulation 1

For our simulations, we used parameters derived from information established before the experiment. One of our goals is to see how our initial model would fare as a substitute for having collected the results of human navigation times. We are ultimately interested in resolving design decisions, which depends on knowing which structures are best and under what conditions. With this aim in mind, we are interested in how well the model qualitatively predicts the human results, and we use the Spearman rank correlation as one metric for how well the model matches the human results. We also present the more traditional Pearson correlation (r), which considers how well the relative distances between the predictions match those of the empirical data.

Parameters

We used the timing parameters from our previous experiments, namely 500 msec for link selecting and pressing the Back button, and 250 msec for evaluating each link. The relevance ratings for each label in the site were derived from the assessments we had already obtained to select the targets. For each pairing of the targeted items (e.g., Birdbath) and the link labels (e.g., Housewares), we and a third judge had rated how likely we thought the given link would lead to the targeted item in a Web site. We chose among three ratings (Probable, Possible but unlikely, and Highly unlikely). By assigning respective values of 1, .5, and 0 and averaging them among three judges, we obtained a range of values from 0 to 1.

For setting selection thresholds, we took the midpoints of the three rating values. Thus, the initial selection threshold was set to .75 and the secondary threshold is set to .25. We use these thresholds so that links assessed as Probable will be selected on the first pass and links assessed as Possible but unlikely will be later selected when the threshold is reduced.

The model evaluated the links in the same order that they were displayed on the participant's browser for each task (recall that these orders were randomized for each task). Sometimes the links required multiple columns. For these cases, the model first evaluated the links in the first column (top to bottom), then the second column, and so on.

Results

Pairing the 24 averaged times of the model's predictions with those from the human results produced a Spearman rank correlation of .739 and a Pearson correlation of .692. Pairing the number of link evaluations performed by the model with the times from the human results produced a rank correlation of .717. The rank correlation using the number of link selections was .523.

Discussion

The Pearson correlation (r) indicates that the model accounts for 47.9% (r^2) of the human performance data. The number of link evaluations is nearly as good as a predictor for the human results as the model's simulated time. If the number of link evaluations generally compares favorably with the simulated time (which relies on time constants), it might ultimately serve as a parameter-free predictor of actual search times.

Although using the mean of judged ratings accounts for nearly half of the variation in the empirical data, we might improve the model's predictions by capturing the variability in the judged ratings. This would certainly be the case if the relation between ambiguity and navigation time were not linear. For example, targets behind exceptionally ambiguous labels may take an amount of time that is an order of magnitude larger than targets behind exceptionally clear labels. In theses cases, ratings with the same mean but different variances would produce different mean navigation times.

In the next simulation, we consider the role of rating variance. To represent variances of how users assess labels, we added noise to the judged ratings in proportion to how inconsistent the judges' ratings were (as measured by standard deviation). When all three judges agreed, no noise is added. From a statistical viewpoint, this method uses judges' assessments to estimate the parameters of a normal distribution that would describe the actual assessments. Although we do not know whether actual relevance ratings fit a normal distribution, this distribution is consistent with the assumption that actual relevance ratings are concentrated around an estimated mean.

5.2. Simulation 2

Parameters

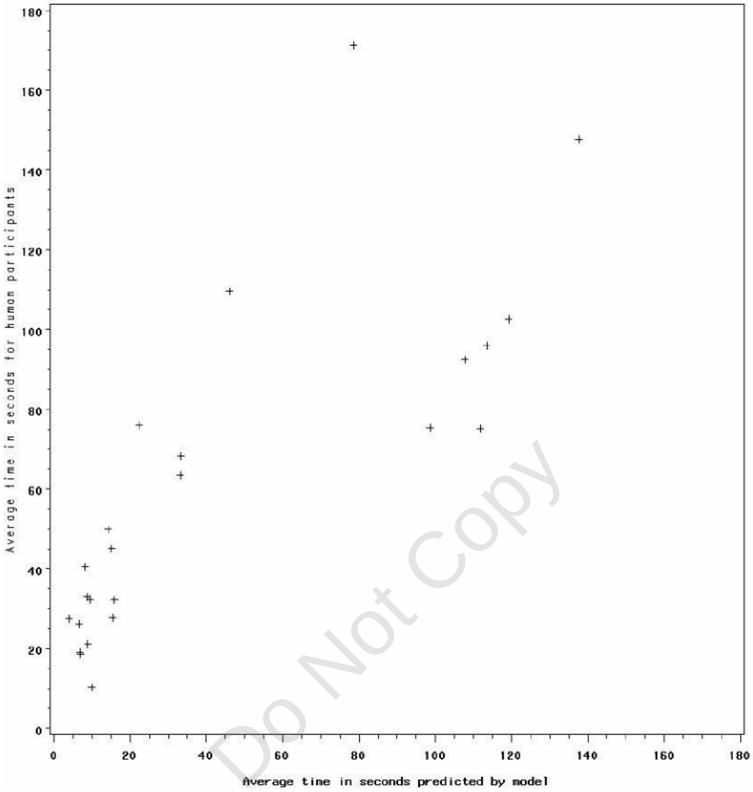
The parameters were the same as those in Simulation 1 except for the inclusion of a random variable for adding variation to the judges' average rating. The random variable followed a normal distribution whose mean and standard deviation were taken from the judges' ratings. Because the model now has a stochastic element to it, we ran the model 100 times for each task and took the average. This improves the consistency of the model's predictions.

Results

The rank correlation was .851 using the model's simulated times and .790 for the Pearson correlation. Using the model's number of link evaluations and selections, the rank correlations were .880 and .377, respectively.

Figure 16 shows a scatter plot of predicted times (x-axis) versus observed times (y-axis) for each of the 24 conditions, averaged over subjects. All but six of the plots approximate a line. These six plots consist of three different targets (the Tripod Grill on the three-tiered and the 37 × 13 structures; the

Figure 16. Plot comparing simulated predictions with times for human participants.



Shower Organizer on the three-tiered and the 6 × 81 structures; and the Garage Remote on the three-tiered and the 6 × 81 structures). Parameter estimates for the regression line in Figure 16 are 29.4 sec (SE = 7.5) for the y intercept and .73 (SE = .12) for the slope.

Discussion

Factoring in the variability of the judged ratings substantially improved the model's predictions. We arrived at a rank correlation of .851 using parameters established without the benefit of the data collected from our experiment involving human participants. The time costs were established from the menu selection studies. The label relevance values and the model's thresholds were derived from the ratings of three judges. This correlation represents the current model's capability for predicting performance trade-offs in the absence of human participants. The high rank correlation indicates how the model is useful for revealing general performance patterns across different structures and different levels of label reliability.

For absolute predictions, the simulated times underpredict the human times if we exclude the six highest prediction times made by the model. One possible direction for improving absolute predictions could involve increasing the time constants whereas also obtaining better estimates for label relevances. Increasing the time constants would improve the absolute fits of all but the six highest prediction times. In explaining the discrepancy of the six highest predictions, it may be the case that MESA's opportunistic strategy incurs a cost greater than that exhibited by human users when they have difficulty finding a target. Alternatively, it is possible that the judged ratings overestimated the difficulty of a few critical labels for some of the targets. Even a slight overestimation of difficulty can cause MESA to incur substantial time costs as it rescans pages. Referring to Figure 16, judgment differences among the high-level categories (e.g., Housewares and Hardware) leading to the Garage Remote and to the Shower Organizer would explain why the model's predictions of these items deviated from the human results for the three-tiered and the 6 × 81 structures but not the 37 × 13 structure. Similarly, judgment differences among the low-level categories (e.g., Camping Accessories and Cooking Gadgets) would explain the discrepancy for the Tripod Grill in the three-tiered and 37 × 13 structures but not the 6 × 81 structure. Later in this article, we further discuss possible directions for improving estimates of label relevance with the goal of achieving better absolute time predictions.

Another possible source of error is the variance among the judged ratings. It may not accurately represent the variance among the participants. Given the greater diversity of the participants, it is likely that their assessments of the link labels varied substantially more than those of the three judges. We explore this possibility in the next set of simulations where we incrementally increased the variance.

5.3. Simulations with Increased Variance

Simulation 2 used the same variances for the label relevances as those from the judges. To better represent a greater diversity of participants, we ran simulations where we incrementally increased the variance.

Parameters

The parameters were the same as those in Simulation 2 except multiple simulations were conducted where the standard deviation of the random variable was incrementally increased by a factor of 0.5, ranging from 1.5 to 3.0.

Figure 17. Simulation results at increased levels of variance.

Variance Factor	Pearson Correlation	Rank Correlation
1.0	.790	.851
1.5	.816	.863
2.0	.841	.863
2.5	.839	.854
3.0	.832	.869

Results

Figure 17 shows the resulting correlations at increasing levels of variance. For the sake of comparison, the table includes the correlations from simulation 2 (i.e., the variance factor equals 1). For all increased levels of variance, the model's predictions more closely corresponded to the participants' times in terms of Pearson correlations and rank correlations. In particular, a factor of 2 produced the best Pearson correlation and nearly the best rank correlation.

Discussion

The results show that increasing the amount of variance in the link label ratings improves the model's predictions and suggest that these increased amounts better account for the diversity of label assessments among the participants. Because the increased variance occurs among the more ambiguous link labels, the added variance improves the model's predictions for the slower navigation times.

As we have already noted, the quality of the link labels is the principal determiner for how quickly people find items in a Web site. To further illustrate this point, we derived a simple measure of link quality to see how it would correlate with navigation times. The link quality measure is a simple average of the judges' ratings for the labels leading to the targeted item. Because the last label is the target itself, its rating is 1. The measure for a two-tiered structure is an average of two values (including the 1 for the target itself) and the measure for a three-tiered structure is an average of three values (including the 1 for the target itself). Comparing the link quality measures to actual navigation times yielded a Pearson correlation of –.763 and a rank correlation of –.840.

This simple measure accounts for over 58% of the variance, suggesting that a simple link quality measure can provide us with a predictor of navigation times that is nearly as good as the model in the second simulation. This reinforces our claim that the quality of link labels is the dominating factor for how quickly people find items in a Web site.

Despite its predictive power, this simple statistical measure has its limitations. For example, it does not consider variances in label assessments, whose value is revealed by increasing the variance factor in our process model. Nor does the statistical measure consider the number or relevance of competing labels. This limitation is best revealed by considering navigation times for targets with ideal link labels. When we only consider the four target-structure searches where participants and the model usually took the optimal path, the correlation for the process model is .967 (.800 for the rank correlation). In contrast, because the link quality measure is a perfect 1 for all four cases, the statistical measure cannot account for any of the variation in their navigation times.

6. GENERAL DISCUSSION

One of our goals was to explore how information structures affect Web site navigation. Previous results from menu selection studies suggested an optimal number of 8 selections per display, whereas results from Larson and Czerwinski's (1998) Web navigation study showed participants finding items faster in structures with 16 and 32 links per page. We account for the discrepancy by showing how the quality of the link labels interacts with the structure of the site. Through the aid of a process model, we showed that two-tiered structures (i.e., 16 × 32 and 32 × 16) produce faster results than a comparable three-tiered structure (i.e., 8 × 8 × 8) under the following conditions:

- The categorical link labels are sufficiently ambiguous so that the user must perform some backtracking to find the item.
- The bottom level consists of labels that clearly indicate which link leads to the targeted item.
- The link labels are not ordered or grouped in a way that would allow a user to confidently skip sets of labels without fully evaluating them.

However, our simulations and empirical study showed that a three-tiered structure produces faster navigation times when the compared structures have clear labels. Our simulations also predict that the three-tiered structure may be optimal when the compared structures have ambiguous labels at all levels. For these cases, when the level of label quality is the same across all levels, our findings are consistent with the theoretical and empirical results from menu selection studies.

These results were achieved by incorporating the following properties into our model MESA:

- Sequential evaluation of labels with a time cost for each evaluation.

- Representing labels at various levels of relevance.
- Modeling the cost of selecting misleading links.

This last property considers cognitive limitations when simulating the expense of returning to the previous page. For example, MESA often needs to rescan a previously visited page because it may not recall the presence of relevant links. This additional cost is substantial for structures with a secondary level containing misleading link labels. However, without misleading links, an additional level, with its fewer links per page, provides a more efficient access to content pages.

The interactive effect that label quality has on choosing optimal structures has implications for research in Web navigation. Our analysis suggests that evaluation methods and empirical studies must consider the quality of the link labels for them to be useful. Evaluation methods that do not consider label quality run the risk of seriously misjudging the quality of a Web site. Similarly, studies that do not account for the quality of link labels may produce misleading results. Future studies may need to separately analyze tasks at varying levels of label quality to produce useful results. Alternatively, it may be possible to manipulate label quality by creating different sets of labels for the same tasks.

We have not considered the effect that grouping or ordering links has on navigation times. Although many Web sites use categories that are not easily grouped or ordered, some Web sites have pages where links are grouped in categories or ordered in a systematic way (such as alphabetically). For these link arrangements, users may be able to effectively skip sets of links to quickly find the link that leads to the targeted item. As a consequence, effectively grouping or ordering links increases the optimal number of links per page (Landauer & Nachbar, 1985; Snowberry et al., 1999).

A common approach for grouping links on a page involves lifting a secondary level of links and placing them under each corresponding label at the upper level. In this way, two levels of the structure appear on one page. There is some evidence that people navigate this within-page structure in the same way that our model navigates a two-level structure across multiple pages. That is, the user scans each header label and, upon finding a relevant label, chooses to scan the secondary labels below it (Hornof & Halverson, 2003). The principal difference is that a structure realized across multiple pages requires a physical user action and system response to select a category label. To the extent that user navigation of structures across pages is similar to that within a page, we could model the navigation of within-page structures by identifying faster time costs for selecting category labels. Alternatively, the model could be used to identify good structures that are initially realized across pages. Later, as the detailed design of the Web site is further developed, the design could be optimized by consolidating levels on one page. This second approach assumes that the best structures realized across multiple pages will be the best structures realized within a page.

Perhaps the most useful lesson for Web designers is the importance of choosing clear and reliable link labels. Our results demonstrate that the quality of link labels is a greater factor for navigation times than the structure of the pages. In our study with human participants, the targets with the best link labels were found faster than those with poor link labels, regardless of the structure. Our simulation results corroborate those findings. Reinforcing this point, the averages of judged label ratings were able to account for 58.2% of the variance in navigation times. We thus advise designers to structure Web sites using the most reliable link labels, rather than trying to achieve pages with the “optimal” number of links if it forces the use of ambiguous labels. For example, a top level with just a half dozen links could be part of an effective structure if the top level links reliably led the user to the next level.

Of course, even the best link labels may not compensate for the additional navigation costs imposed by a structure with an extreme number of links per page. A structure that has a reasonable number of links per page but causes an occasional selection error may still serve the user better. In the absence of any simple guidelines for weighing reasonable structures against ideal link labels, a Web designer may still need to test a variety of structures.

Our second goal was to demonstrate how a process model might be used to test information structures during the design process. Experimenting with a range of information structures with human participants is costly and usually not feasible. Relying on the ratings of three judges and previously established parameters, MESA was able to achieve a rank correlation of .85 when compared to the results collected from human participants. By adding more variance to the label ratings, the model obtained a rank correlation of .863. At this level, the Pearson correlation was .841 and thus accounted for 70.7% of the variance in the empirical data.

We have yet to fully explore alternate methods for estimating label relevance. Some methods are likely to be more accurate but also more costly. For example, one could survey potential users to collect their relevance ratings and apply the same method we used for our judges’ ratings. Other methods may be less accurate but also less costly. For example, many Web sites may have similar distributions of label quality. If true, we could model structural trade-offs on these Web sites by imposing a “typical” distribution of label quality. We might also find that a small sample of judged ratings can provide a useful approximation for typical relevance values throughout the site.

We would also like to consider automatic methods for determining label relevance. We have discussed some efforts that produce a similarity metric between a pair of words or phrases based on their co-occurrences in textual corpora. At this time, we do not know how well these methods produce useful relevance estimates between a target phrase and a link label. Experimenting with a variety of methods, corpora, and navigation domains will help us understand the role of these methods in predicting navigation costs.

Although the model's performance critically depends on the accuracy of the label ratings, additional improvements may also come by better understanding how people scan, evaluate, and select links. For example, MESA's scanning strategy makes the simplifying assumption that people require a fixed time to evaluate each link label. In reality, people require varying amounts of time that probably increase when a label's relevance is close to the selection criterion. We may also find that sometimes people use a comparison strategy instead of the threshold strategy we used in our simulations. As we obtain better estimates of link relevance, we will be able to explore alternate methods such as these refinements and learn which of them produce more accurate predictions.

NOTES

Acknowledgments. We thank Susan Dumais, Peter Pirolli, and two anonymous reviewers for their helpful comments on earlier versions of this article. We acknowledge Gillian Jetson who collected the department store categories and items from a local store.

Support. Funding for this work was provided by the Human-Automation Integration Research portion of the Airspace Operation Systems Project of NASA's Airspace Systems Program.

Authors' Present Addresses. Craig S. Miller, DePaul University, School of CTI, 243 S. Wabash Avenue, Chicago, IL 60604-2302. E-mail: cmiller@cs.depaul.edu. Roger W. Remington, MS 262-4, NASA Ames Research Center, Moffett Field, CA 94035. E-mail: Roger.W.Remington@nasa.gov.

HCI Editorial Record. First manuscript received February 17, 2003. Revision received January 6, 2004. Accepted by Susan Dumais. Final manuscript received April 7, 2004. —Editor

REFERENCES

Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bernard, M. L. (2002). Examining a metric for predicting the accessibility of information within hypertext structures. Unpublished doctoral dissertation, Wichita State University.

Blackmon, M. H., Kitajima, M., & Polson, P. G. (2003). Repairing usability problems identified by the cognitive walkthrough for the Web. Proceedings of the Conference on Human Factors in Computing Systems (pp. 497–504). New York: ACM.

Blackmon, M. H., Polson, P. G., Kitajima, M., & Lewis, C. (2002). Cognitive walkthrough for the Web. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 463–470). New York: ACM.

Broadbent, D. E. (1958). Perception and communication. New York: Pergamon.

Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. (1999). The tangled web we wove: A taskonomy of WWW use. Proceedings of CHI' 99 Human Factors in Computing Systems (pp. 544–551). New York: ACM.

Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human–computer interaction. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information needs and actions on the Web. Proceedings of the CHI 2001 Conference on Human Factors in Computing Systems (pp. 490–497). New York: ACM.

Chi, E. H., Rosien, A., Supattanasiri, G., Williams, A., Royer, C., Chow, C., et al. (2003). The bloodhound project: Automating discovery of Web usability issues using the infoscent simulator. Proceedings of the Conference on Human Factors in Computing Systems (pp. 505–512). New York: ACM.

Furnas, G. W. (1997). Effective view navigation. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 367–374). New York: ACM.

Hornof, A. J., & Halverson, T. (2003). Cognitive strategies and eye movements for searching hierarchical computer displays. Proceedings of the Conference on Human Factors in Computing Systems (pp. 249–256). New York: ACM.

Johnston, J. C., McCann, R. S., & Remington, R. W. (1995). Chronometric evidence for two types of attention. Psychological Science, 6, 365–369.

Kiger, J. I. (1984). The depth/breadth trade-off in the design of menu-driven user interfaces. International Journal of Man-Machine Studies, 20, 201–213.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, 104, 211–240.

Landauer, T. K., & Nachbar, D. W. (1985). Selection from alphabetic and numeric menu trees using a touch screen: Breadth, depth, and width. Proceedings of CHI' 85 Human Factors in Computing Systems (pp. 73–78). New York: ACM.

Larson, K., & Czerwinski, M. (1998). Web page design: Implications of memory, structure and scent for information retrieval. CHI' 98: Human Factors in Computing Systems (pp. 25–32). New York: ACM.

Lee, E., & MacGregor, J. (1985). Minimizing user search time in menu retrieval systems. Human Factors, 27, 157–162.

Lynch, G., Palmiter, S., & Tilt, C. (1999). The max model: A standard Web site user model. Proceedings of the 5th Annual Human Factors and the Web Conference. Retrieved August 26, 2002 from http://zing.ncsl.nist.gov/hfweb/proceedings/ lynch/index.html.

McCann, R. S., Folk, C. L., & Johnston, J. C. (1992). The role of attention in visual word processing. Journal of Experimental Psychology: Human Perception and Performance, 18, 1015–1029.

Miller, C. S., & Remington, R. W. (2000). A computational model of Web navigation: Exploring interactions between hierarchical depth and link ambiguity. The 6th Conference on Human Factors & the Web. Retrieved August 26, 2002 from http://www.tri.sbc.com/ hfweb/ miller/article.html.

Miller, C. S., & Remington, R. W. (2001). Modeling an opportunistic strategy for information navigation. Proceedings of the Twenty-Third Conference of the Cognitive Science Society (pp. 639–644). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Miller, C. S., & Remington, R. W. (2002). Effects of structure and label ambiguity on information navigation. Conference Extended Abstracts on Human Factors in Computer Systems (pp. 630–631). New York: ACM.

Miller, D. P. (1981). The depth/breadth tradeoff in hierarchical computer menus. Proceedings of the Human Factors Society, Vol. 25 (pp. 296–300). Santa Monica, CA: The Human Factors and Ergonomics Society.

Neisser, U. (1967). Cognitive psychology. New York: Appleton-Century-Crofts.

Norman, K. L. (1991). The psychology of menu selection: Designing cognitive control at the human/computer interface. Norwood, NJ: Ablex.

Peck, V. A., & John, B. E. (1992). Browser soar: A computational model of a highly interactive task. Proceedings of CHI'92 Human Factors in Computing Systems (pp. 165– 172). New York: ACM.

Pirolli, P., & Card, S. (1999). Information foraging. Psychological Review, 106, 643–675.

Pirolli, P., & Fu, W.-T. (2003). SNIF-ACT: A model of information foraging on the World Wide Web. In J. G. Carbonell & J. Siekmann (Eds.), Lecture Notes in Artificial Intelligence, 27(2). Heidelberg: Springer-Verlag.

Posner, M. E. (1980). Orienting of attention. Quarterly Journal of Experimental Psychology, 32, 3–25.

Rosenfeld, R., & Morville, P. (1998). Information architecture for the World Wide Web. Sebastopol, CA: O'Reilly & Associates.

Shneiderman, B. (1998). Designing the user interface: Strategies for effective human–computer interaction, third edition. Reading, MA: Addison-Wesley.

Snowberry, K., Parkinson, S. R., & Sisson, N. (1983). Computer display menus. Ergonomics, 26, 699–712.

Soto, R. (1999). Learning and performing by exploration: Label quality measured by latent semantic analysis. Proceedings of CHI' 99 Human Factors in Computing Systems (pp. 418–425). New York: ACM.

Sperling, G. (1960). The information available in brief visual presentations. Psychological Monographs, 74(11, Whole No. 498).

Young, R. M. (1998). Rational analysis of exploratory choice. In M. Oaksford & N. Chater (Eds.), Rational models of cognition (pp. 469–500). Oxford, UK: Oxford University Press.

Small Electronics
Bread Bakers
Bread slicing guide
Automatic bread maker
Bread maker recipe book
Bread slicing system
Bread machine mix

Coffee Makers
Coffee cappuccino machine
Coffee maker with timer
Replacement coffee filters
Coffee maker cleaner
Space maker coffee maker
Espresso/cappuccino maker

  - Water filter
  - Coffee filter basket
  - Thermal replacement carafe
  - Replacement carafe
  - Coffee grinder/espresso mill
  - Cone coffee filters
  - Auto party perk (coffee urn)
  - Espresso machine
  - Gold coffee filter
  - Replacement decanter
  - Coffee grinder
  - Permanent coffee filters
  - Stainless steel coffee maker
- Tea Makers
  - Electric teakettle
  - Cordless electric kettle
  - Hot tea maker
  - Iced teapot
  - Hot pot
  - Teakettle
- Crock Pots/Slow Cookers
  - Crock pot
  - Electric deep fryer
  - Electric fry pan
  - Fondue pot
  - Rice cooker
  - Sandwich maker
  - Waffle maker
  - Electric skillet
  - Roaster oven
  - Electric kitchen kettle (multi cooker/steamer)
  - Electric wok
  - Food steamer
- Indoor Grills
  - Carousel rotisserie
  - Space-saving buffet range
  - Cool-touch griddle
  - Indoor electric grill
  - Grill machine
  - Dual burner buffet range
  - Indoor tabletop grill
  - Beef jerky works gun
  - Beef jerky spices
  - Electric griddle
- Hand/Stand Mixers
  - Rock salt for homemade ice cream
  - Spatula mixer
  - Ice cream mix
  - Hand blender with chopper and disk
  - Electric ice cream maker
  - Hand mixer
  - Auto ice cream maker
  - Hand/stand mixer
  - Stand mixer
  - Ice cream and yogurt freezer
  - Hand blender
  - Handy chopper

- Hardware
  - Home Hardware
    - Barrel hold
    - Picture hanger
    - Cup hook
    - Doorstop
    - Hook and eye
    - Over door hook
    - Storage hook
    - Garment hook
    - Towel grips
    - Plant brackets
  - Electrical Supplies
    - Polarized cube taps
    - Wall hugger tap
    - Fuse
    - Dimmer knob
    - Color tape
    - Grounding outlet
    - Plug fuse
    - Lighted dimmer knob
    - Surge protector
    - Cable ties
    - Rotate-on dimmer knob
    - Multiple outlet center
    - Safety caps
    - Power strip
    - Push-on dimmer knob
  - Hand Tools
    - Long nose pliers
    - Socket set
    - Wrench
    - Clamp
    - Adjustable pliers
    - Bit driver set
    - Hex key set
    - Scraper
    - Hacksaw
    - Hammer
    - Screwdriver

- Home Security
  - Door guard
  - Smoke and carbon monoxide detector
  - Fire escape ladder
  - Surface bolt
  - Timer
  - Lamp appliance security timer
  - Security floodlight
  - Carbon monoxide alarm
  - Door viewer
  - Fire extinguisher
  - Garage remote
  - Automatic light
- Bath Storage
  - Etagere
  - Paper holder
  - 3-shelf pole caddy
  - Wastebasket
  - Towel bar
  - Shower basket
  - Shower caddy
  - Storage tower
  - Shower organizer
  - Bath caddy
  - Robe hook
  - Towel ring
- Decorative Lighting
  - Black light
  - High intensity bulb
  - Black party light
  - Indoor spotlight
  - Nite light
  - Red party light
  - Pink party light
  - White fan bulb
  - White blunt tip bulb
  - Clear globe
  - Purple party light
  - Tubular bulb
  - Clear blunt tip bulb
  - White flame tip bulb
  - White globe light
  - Green party light
- Plumbing
  - Snap-on aerator
  - Drain strainer
  - Toilet seat bolts
  - Basket strainer
  - Toilet flapper
  - Tub sealer
  - Sprayer head
  - Tile trim
  - Basin stopper
  - Toilet seat hinges
  - Toilet bolt caps
  - Tub stopper

- Automotive
  - Tire Care
    - Tire glaze
    - Tire foam shine
    - Wheel cleaner
    - Tire wet
    - Tire scrubber
    - Tire shine
    - Tire max
    - Bleach white
    - Tire care
    - Wheel detail brush
  - Cell Phone Accessories
    - Rapid charger
    - Leather case
    - Speakeasy headset
    - Boom mike
    - Cellular passive repeater antenna
    - Dash mount phone holder
    - Sport phone case
    - Phone holder
  - Interior Care (Automotive)
    - Amorall interior
    - Quick detailer
    - Leather cleaner and conditioner
    - Fabric refresher
    - Carpet and upholstery cleaner
    - Odor eliminator
    - Power brush vacuum
    - Foam cleaner
    - Spot and stain remover
    - Dash duster
  - Automotive Safety
    - Safe lights
    - Laminated steel padlock
    - The Club®
    - Emergency road kit
    - First aid kit
    - Padlocks
    - Strobe light
    - Alarm

  - Auto Accessories
    - Backseat organizer
    - Fleece seat belt caddy
    - Visor organizer
    - Seat belt shoulder pad
    - CD visor organizer
    - Nylon shoulder pad
    - Fleece car seat organizer
    - Litterbag
  - Oil
    - Engine treatment
    - Oil funnel
    - Oil drain
    - Smoke treatment
    - Engine degreaser
    - Oil filter
    - Stop leaks
    - Fuel system cleaner
    - Oil
    - Fuel injection treatment
    - Oil treatment
  - Car Wash
    - Sponge
    - Bug-gone scrubber
    - Wash mitt
    - Chamois squeegee
    - Chamois
    - Bucket
    - Scrubbing pad
    - Dip and wash brush
    - Leather dryer
    - Shampoo wash pad
    - Wash pad
    - Vehicle wash brush

- Houseware
  - Cooking Utensils
    - Baking spatula
    - Spatula
    - Balloon whisk
    - Pasta fork
    - Slotted ladle
    - Stir-fry scoop
    - Fork
    - Slot spoon
    - Measuring spoon
    - Ladle
    - Tongs
    - Pastry brush
    - Measuring cups
    - Spoon
    - Slot spatula
  - Glass Drinkware
    - Margarita glass
    - Bouquet wine glass
    - Flute
    - Mug
    - Juice glass
    - Iced tea glass
    - Tumbler
    - Cooler
    - Martini glass
    - Goblet
    - Wine glass
    - Shot glass
    - Cognac glass
  - Dinnerware
    - Canister
    - Spoon holder
    - Dessert plate
    - Salad bowl
    - Bowl
    - Popcorn bowl
    - Dinner plate
    - Serving bowl
    - Cups and saucers
    - Oval platter
    - Rectangular baker
    - Pepper shaker
    - Salt shaker
    - Mugs
    - Soup bowl
    - Round platter
    - Square plate
    - Ice cream bowl
    - Salad bowl
    - Pasta bowl
  - Cookware
    - Round grill pan
    - Round pan
    - Loaf pan
    - Pourable saucepan lids
    - Saucepan
    - Bake pan
    - Open saucepan
    - Bake pan
    - 10-piece cookware set
    - Square pan
    - Universal steamer insert
    - Jelly roll pan

  - Sauté pan
  - Open skillet
  - Muffin pan
  - Omelet pan
  - Griddle
  - Stir fry pan
  - Cookie sheet
  - Round cake pan
- Specialty Cooking
  - Salsa bowl and ladle
  - Bread warmer basket
  - Tortilla griddle
  - Pizza peel
  - Bread baking stone
  - Chopsticks
  - Pizza baking stone and rack
  - Mexican griddle
  - Stir-fry cookbook
  - Everyday fiesta cookbook
  - Bread baking stone and rack
  - Tortilla warmer
  - Wok set
  - Pizza and pasta cookbook
  - Wok
  - Fiesta taco holders
- BBQ Tools and Gadgets
  - Grill brush
  - Can opener
  - Steel tongs
  - Grill basket
  - Salt shaker
  - Turner
  - Sugar shaker
  - Grill basket
  - Corn holders
  - BBQ skewers
  - Basting brush
  - BBQ set
  - BBQ tongs
  - BBQ brush
  - Thermometer
  - Salad tongs
- Kitchen Gadgets
  - Utility hooks
  - Bag clip
  - Splatter guard
  - Wooden plate easel
  - Small bowl
  - Faucet nozzle
  - Sugar holder
  - Toothpick holder
  - Sugar pourer
  - Ashtray
  - Salt and pepper shakers
  - Plate cover
  - Grater
  - Chip clips
  - Magnetic clips
  - Wedger/corer
  - Shakers
  - Cheese shaker
  - Strainer
  - Sink stopper

Garden
- Grill Accessories
  - Mitt
  - Porcelain grill
  - Steak basket
  - Heavy-duty matches
  - Basic kettle cover
  - Grill scrubber
  - Kabob set and frame
  - BBQ set
  - Tool holder
  - Spatula
  - Stainless steel forks
  - Charcoal lighter
  - Electric rotisserie
  - Basting set
  - Tongs
  - Butane lighter
  - Thermo fork
  - Charcoal
  - Grill brush
- Patio Accessories
  - Outdoor clock
  - Mud brush
  - Thermometer
  - Stepping stone
  - Sandstone candleholder
  - Hose guide
  - Gazing ball metal stand
  - Wall plaque
  - Gazing ball base
  - Birdbath
  - Gazing ball
- Patio Furniture
  - Wrought iron chair
  - Cushioned swing

- Folding chair
- Clamp with umbrella
- White steel accent table
- Children's sand chair
- Captain's chair
- Resin adirondack
- Lounger
- Resin chair
- Chair
- Resin table

- Sporting Goods
  - Lanterns
    - Dynalight
    - Flashlight
    - Table lamp
    - Mantles
    - Lantern spark igniter
    - Headlight
    - Tent light
    - Lantern
    - Emergency candles
    - Area light/flashlight
    - Rechargeable lantern
    - Sport light
    - Propane lantern
    - Floating lantern
    - Tub candles
    - Candle lantern
    - Replacement globe
  - Knives/Multi Tools/Two-way Radios
    - 2-way radio
    - Flashlight/radio
    - Camper's tool
    - Pocket sharpener
    - Swisscard
    - Serrated knife
    - Multilock
    - Pocketknife
    - Walkie-talkies
    - Multiplier
    - Swiss Army knife
  - Weights/Fitness/Exercise
    - Sport towel
    - Slimmer short
    - Mesh gloves
    - Heavy tension spring grips
    - Lycra gloves
    - Wrist ring
    - Wrist band
    - Dumbbell set
    - Waist slimmer belt
    - Walking weights
    - Handheld weights
    - Wrist/ankle weights
    - Headband
    - Resistance band
    - Contour belt
    - Cast iron hex dumbbells
    - Squeeze ball
    - Neoprene fitbell
    - Sauna suit
  - Tents
    - Tent peg mallet
    - Braided polyester cord
    - ABS tent stakes
    - Steel tent pegs
    - Tent whisk and dustpan
    - Canvas tent repair kit
    - Heavy-duty tarpaulin
    - Tent stake puller
    - ABS tent pegs
    - Guy ropes and slides
  - Fishing Rods/Fishing Reels
    - Fishing line
    - Spinner bait
    - Fishhooks
    - Protective eyeglasses
    - Brass snap swivels
    - Horizontal rod rack
    - Dip net
    - Spinning reel
    - Stringer
    - Minnow spin
    - Fishing rod
    - Bobber stops
    - Jighead
    - Float assortment
    - Brass casting sinkers
    - Power bait
    - Maribu jigs
  - Bike Accessories
    - Seat wedge pack
    - Water bottle
    - Frame bag
    - Handlebar water bottle
    - Handlebar bag
    - Bike seat
    - Kickstand
    - Seat cover

  - Bike glove
  - Headlight / mirror/bell set
- Camping Accessories
  - 3-piece knife/fork/spoon set
  - Cast iron griddle
  - Propane
  - Tripod grill
  - Nesting utensil set
  - Plastic cutlery set
  - Enamel plate
  - Grill pan
  - Griddle
  - Charcoal water smoker and grill
  - Extendable cooking fork
  - Nylon spoon
  - Nylon spatula
  - Cast iron Dutch oven
  - Enamel percolator
  - 5-piece mess kit
  - Hand grill
  - Enamel kettle with cover
  - Enamel bowl
  - Enamel mug

APPENDIX

The following is the site structure used for the empirical study and the detailed simulations. The order of the items was randomized in the study and is thus arbitrary.